The use of ocean reanalysis products to initialize ENSO predictions

Youmin Tang,[1] Richard Kleeman,[1] Andrew M. Moore,[2] Anthony Weaver,[3] and Jérôme Vialard[4]

Received 7 May 2002; accepted 11 June 2003; published 9 July 2003.

[1] With a three-dimensional variational (3D-Var) assimilation scheme and a hybrid coupled model, we have explored the possibility of initializing ENSO prediction models by assimilating NCEP (National Centers of Environment Prediction) reanalysis subsurface temperatures. Our results show that, compared to predictions without assimilation, the reanalysis product can effectively improve prediction of both Niño3 sea surface temperature anomalies (SSTA) at all lead times up to 12 months (in particular for lead times over 4–6 months) and of El Niño episodes. The oceanic analysis from the assimilation with the reanalysis product can be as good as those generated by directly assimilating subsurface in situ temperature observations. *INDEX TERMS:* 4522 Oceanography: Physical: El Niño; 4504 Oceanography: Physical: Air/sea interactions (0312); 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 4263 Oceanography: General: Ocean prediction. **Citation:** Tang, Y., R. Kleeman, A. M. Moore, A. Weaver, and J. Vialard, The use of ocean reanalysis products to initialize ENSO predictions, *Geophys. Res. Lett.*, *30*(13), 1694, doi:10.1029/2003GL017664, 2003.

1. Introduction

[2] It has been found that the initialization with subsurface in situ temperature observations can significantly improve ENSO prediction skills [e.g., *Ji et al.*, 1998]. However even though the subsurface in situ observations are relatively spatially sparse and temporally sporadic [*McPhaden et al.*, 1998], a great deal of effort is required to process and assimilate the data in models. The assimilation of this type of data is therefore often confined to specialized assimilation groups or a few national operational forecast centers with well-designed assimilation algorithms and high resolution models.

[3] Using a relatively simple but effective initialization scheme is of interest to many ENSO researchers. Typical examples are the predictability studies, which require large numbers of initial condition for ensemble predictions. It is therefore of interest to ask whether it is possible to initialize ENSO prediction models with a reanalysis product? Compared with sparse and sporadic observations, reanalysis products are easier and more convenient to use, since they are usually regular gridded datasets. The high resolution of reanalysis products also allows us to use them for all types of oceanic models. This is different to subsurface in situ observations which are often difficult to use in intermediate models with low resolution. In addition, reanalysis products can provide us with initialization data for many variables which are not directly observed.

[4] There exist two essential issues when applying a reanalysis product to initialize ENSO prediction. The first one is whether a reanalysis product generated by an ocean model is consistent with another ocean model which will be initialized? The model biases from the model generating the reanalysis product might have a large influence on the model being initialized. The second issue is how much observed information is lost when a reanalysis product is used?

[5] *Syu and Neelin* [2000] simply inserted NCEP reanalysis subsurface temperature into their OGCM, called as piggyback scheme, and found an improvement in ENSO prediction skill with the initialization scheme. However in *Syu and Neelin* [2000] the OGCM and forcing fields were very similar to those used in the NCEP assimilation system. In this note, we explored more generally the possibility of using a reanalysis product to initialize an oceanic general circulation model (OGCM) for ENSO prediction. The NCEP reanalysis subsurface temperature (*Behringer et al.*, 1998; referred to as NCEP data hereafter) was used because (i) it is promptly updated each month so that routine prediction using this product is possible; and (ii) NCEP used GFDL Modular Ocean Model (MOM2) for its assimilation system, which is rather different from our OGCM, and both models use different forcing fields. These enable the results reported in this note to be of general applicability. This paper is structured as follows: Section 2 briefly describes the coupled model and the initialization scheme. Section 3 examines the initialization scheme by way of hindcast experiments. Section 4 compares the oceanic analyses using NCEP data with those generated by in situ temperature observations, followed by a short summary and discussion in Section 5.

2. The Coupled Model and Assimilation Scheme

[6] A hybrid coupled model (HCM), an OGCM coupled to a statistical atmosphere, was used to explore and test our initialization scheme. The ocean model used is based on OPA version 8.1 [*Madec et al.*, 1999], a primitive equation OGCM. The model uses an Arakawa *C* grid, and was configured for the tropical Pacific ocean between 30°N–30°S and 120°E–75°W. The horizontal resolution in the zonal direction is 1°, while the resolution in the meridional direction is 0.5° within 5° of the equator, smoothly increasing to 2.0° at 30°N and 30°S. There are 25 vertical levels

[1]Courant Institute of Mathematical Sciences, New York University, New York, New York, USA.
[2]Program for Atmosphere Ocean Science, University of Colorado, Boulder, USA.
[3]Climate Modeling and Global Change Group, C.E.R.F.A.C.S., Toulouse Cedex, France.
[4]LODYC, Paris, France.

0094-8276/03/2003GL017664$05.00

with 17 concentrated in the top 250 m of the ocean. The time step for integration was 1.5 hours. The boundaries were closed, with no slip conditions. The model was spun up with monthly observed wind stress and heat flux Q_s as forcing fields, where Q_s was represented by climatological heat flux Q_0, obtained from the ECMWF reanalysis project, plus a relaxation term to T_0, the observed climatological SST. The detailed formulation of this configuration of the ocean model is described in *Vialard et al.* [2002].

[7] The atmospheric model is a linear statistical model identical to *Barnett et al.* [1993], which predicts the contemporaneous surface wind stress field from the surface sea temperature anomalies. The seasonal variations of the responses of winds tress to SST were included in the construction of atmospheric model; i.e., for each month there is essentially a different atmospheric model. The model was trained with NCEP atmospheric reanalysis wind products and Reynolds-Smith SST observations [*Smith et al.*, 1996] from 1951–1980. Therefore, the hindcast experiments carried out from 1989–1998 in the next section are completely independent of the construction of the atmospheric model. This strategy will remove any artificial skill when evaluating the hindcast experiments.

[8] The data assimilation system is the 3-dimensional variational (3D Var) assimilation method described by *Derber and Rosati* [1989]. In the 3D-Var approach, observations are combined with a model background (guess) field by minimizing a cost function that measures the statistically weighted differences between the model state and these two sources of information (see *Derber and Rosati* [1989] and *Tang et al.* [2003] for further details). The 3D-Var method is used here to combine output from the NCEP ocean analysis (hereafter referred to as "NCEP data" for convenience) with the model background fields. NCEP data from the upper 17 levels, ranging from 5 m to 240 m, was used. NCEP data was vertically interpolated into each model layer with a simple linear algorithm prior to assimilation.

[9] As the 'observations' used here are actually from a reanalysis, we adopt the simple strategy of thinning the reanalysis 'observations' to reduce the effects of correlated observation errors, i.e, discarding 'data' at grid points on even-order latitudes. The observation error variances are set to $(0.5°C)^2$ in this study.

3. ENSO Prediction

[10] Now we examine the predictions of the HCM described above. Three initialization schemes, one from the control run (simply forced with observed wind stress), another one from the assimilation scheme forced with the same observed wind stress, and the last one from Syu and Neelin's piggyback scheme [*Syu and Neelin*, 2000], were used. A total of 120 hindcast cases of 12 months duration were made for each experiment from January 1989 to December 1998, starting at the beginning of each month (e.g., 1 January, February 1…December 1).

[11] Figure 1a shows correlation skills of the predictions initialized by the control run and NCEP data, where the predicted Niño3 SSTA is compared against the observed values. Obviously, the prediction skills initialized by NCEP data beat those initialized from the control run, especially for lead times greater than 4–6 months. The prediction skills

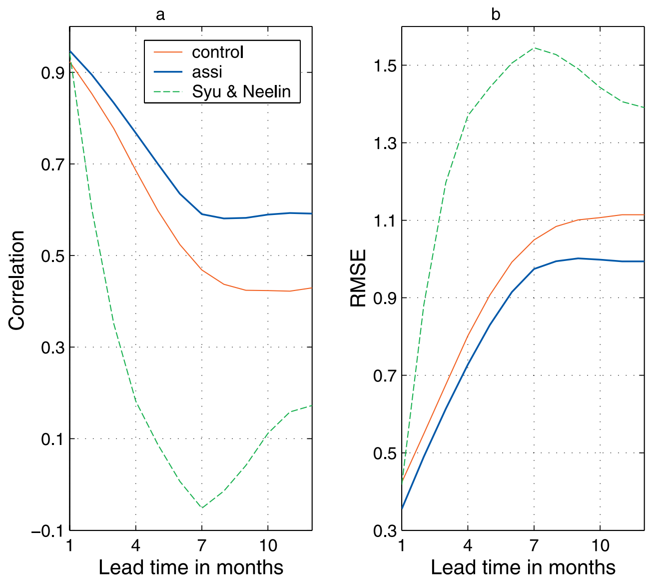

Figure 1. (a) Correlation and (b) RMSE between observed and predicted SST anomalies in the Niño3 region, as a function of lead time. The predictions are initialized every month from January 1989–December 1998. Blue line is from the prediction initialized from NCEP subsurface temperature. Red line is from the prediction initialized from the control run. The green line is from the prediction initialized from Syu and Neelin's piggyback scheme.

from the two schemes display similar characteristics, i.e., the skills degrade with lead time in the first few months and then stabilize until the end. But with the initialization by NCEP data, the rate of decrease of skill is slower, and the period of the decrease is 2–3 months shorter. Similarly, there are also improvements in RMSE (root mean square of error) of the predictions initialized by NCEP data (Figure 1b).

[12] As can be seen in Figure 1, the piggyback scheme leads to a much worse skill than our assimilation scheme, even worse than the control run. This is probably because the piggyback scheme introduces considerable inconsistency into the model, leading to the imbalance of the model variables. Compared with our assimilation scheme, the piggyback scheme only requires the interpolation of NCEP data to the model grid and there is no optimization involved. In fact, the piggyback scheme is a specific case of our assimilation scheme. When the weights to the model background are set to zero (i.e., the reanalysis data is assumed to provide all useful information), our assimilation scheme simply becomes the piggyback scheme. This assumption might approximately hold if the model and forcing that are to be initialized are very similar to those which generated the reanalysis dataset (as in *Syu and Neelin,* 2000). However, as the model and forcing used for the reanalysis dataset are rather different from that which are to be initialized, the reanalysis data certainly fails to provide all useful information, and the model background must be included.

[13] Figure 2 shows the predicted SSTA along the equator at the lead time of 12 months using the two initialization schemes. It was found that the initialization from NCEP data led to better predictions for ENSO warm events, especially with regard to the strength and phase. Generally

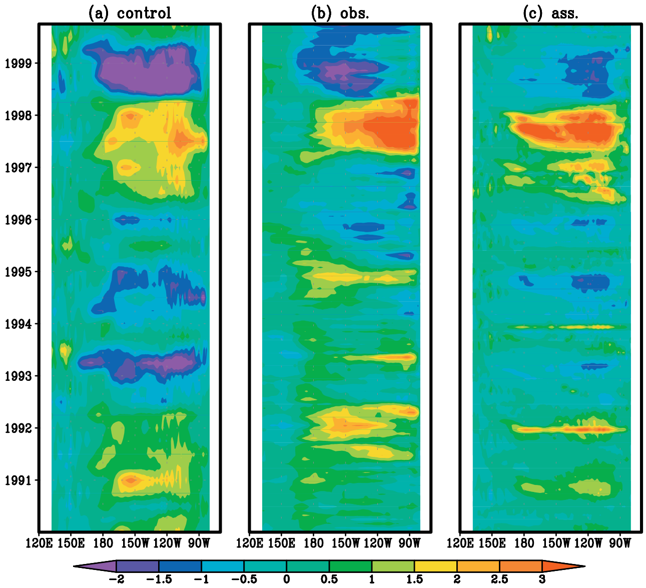

Figure 2. Time-longitude diagrams of predicted SSTA with HCM along the equator at the leadtime of 12 months, initialized by the control run (a) and by NCEP data (c). For comparison, the observed SSTA is shown (b).

there were some spurious phase precedences and relatively weak amplitudes in El Niño predictions initialized for the control run such as the predictions for 1991/1992 and 1997/1998. These deficiencies, however, were alleviated considerably in the predictions initialized by NCEP data as shown in Figure 2c.

4. Comparisons of Different Ocean Analyses for Initialization

[14] An issue inherent in the assimilation of reanalysis products are potential inconsistencies between the reanalysis products and the ocean model into which the reanalysis is assimilated. These inconsistencies are mainly caused by model errors which can lead to some spurious features in reanalysis products. When assimilated into a different model, these spurious features can result an imbalance between the model variables. In addition, reanalysis products no longer contain all of the information from the observations since they are generated by a numerical model, which can either filter observation noise and keep useful information or degrade useful information. The fact that the prediction skills initialized by NCEP data are much better than these initialized by the control run suggests that neither of the above factors is significant in this case. In this section, we will further explore these issues by comparing our oceanic analyses generated by assimilating NCEP reanalysis with those generated by a univariate 3D-Var system [*Weaver et al.*, 2003, *Vialard et al.*, 2003] that has been developed for the same OGCM used in the coupled model of this study. Their 3D-Var scheme assimilates subsurface in situ temperature observations from the Global Temperature and Salinity Pilot Programme dataset, and is quite similar to the one used here. Ocean analyses were produced with that system for the period 1993–98 so the following discussion is confined to that period.

[15] The first variable examined is SST, since it is a direct predictant and represents significant ENSO characteristics. Shown in Figure 3a is the correlation between SST analysis of NCEP data and the 3D-Var SST analysis of *Weaver et al.* [2003] and *Vialard et al.* [2003], indicating that both analyzed values are in very good agreement each other. The good agreement between the two analyzed SSTs can also be seen in their RMSE map and in a Hovmöller diagrams of SSTA along equator (not shown). The high correlation also can be found between either SSTA analysis against observed SSTA (not shown).

[16] Another important component of the coupled ocean-atmosphere system on the interannual timescale is the upper ocean heat content (HC), which is the source of memory for the system. Assimilation of NCEP data results in HC simulations comparable to those from the assimilation of in situ observation (Figure 3b). HC anomalies from both analyses show a better agreement with observations than the control run, in particular in the phase simulation (not shown).

5. Discussion and Summary

[17] Generally, there is a practical problem afflicting the initialization of an OGCM for ENSO prediction: the difficulty of processing and applying useful observations, because subsurface in situ temperature observations are sparse and sporadic. The problem might not be intractable theoretically, but the solution often is time-consuming and very costly. In this note, we have explored the possibility of using NCEP data to initialize an OGCM for ENSO prediction. Using a hybrid coupled model and hindcast experiments, we have presented evidence that initialization by NCEP data can result in a significant improvement in ENSO prediction skill for lead times over 4–6 months. The oceanic analyses generated by NCEP data are as good as those created from in situ temperature observations. This suggests that costly data assimilation using subsurface in situ temperature observations is not required for all ENSO prediction systems. Alternatively, a relatively simple assimilation method using NCEP data might be a more efficient choice.

[18] It should be noted that a reanalysis product like NCEP data was also originally generated by subsurface in situ observations, so subsurface in situ observations are still of great importance. However, this note proposes and validates a relatively simple and effective initialization strategy for ENSO predictions. In this sense, it has practical significance.

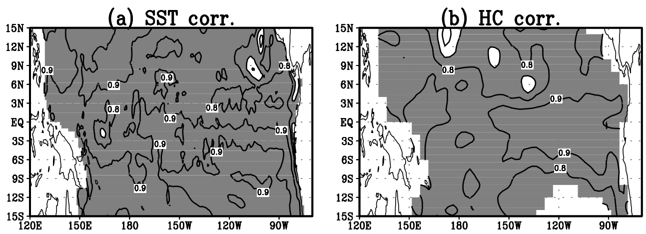

Figure 3. Correlation of SST (a) and HC (b) between the analysis value generated by NCEP data and by in situ temperature observation. The contour interval is 0.1 and the correlation over 0.7 is shaded.

[19] **Acknowledgments.** This work is supported by NSF grant 25-74200-F0960.

References

Barnett, T. P., M. Latif, N. E. Graham, M. Flügel, S. Pazan, and W. White, ENSO and ENSO related predictability, Part1: Prediction of equatorial sea surface temperature with a hybrid coupled ocean-atmosphere model, *J. Climate*, *6*, 1545–1566, 1993.

Derber, J., and A. Rosati, A global oceanic data assimilation system, *J. Phys. Oceanogr.*, *19*, 1333–1347, 1989.

Behringer, D. M., M. Ji, and A. Leetmaa, An improved coupled model for ENSO prediction and implications for ocean initialization -PART I- The ocean data assimilation system, *Mon. Wea. Rev.*, *126*, 1013–1021, 1998.

Ji, M., D. W. Behringer, and A. Leetmaa, An improved coupled model for ENSO prediction and implications for ocean initialization -PART II- The coupled model, *Mon. Wea. Rev.*, *126*, 1022–1034, 1998.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, OPA 8.1 Ocean General circulation model reference manual, Institut Pierre Simon Laplace (IPSL), 91 pp., 1999.

McPhaden, M. J., et al., The Tropical Ocean Global Atmosphere (TOGA) observing system: A decade of progress, *J. Geophys. Res.*, *103*(C7), 14,169–14,240, 1998.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, Reconstruction of historical sea surface temperatures using empirical orthogonal functions, *J. Climate*, *9*, 1403–1420, 1996.

Syu, H.-H., and D. Neelin, ENSO in a hybrid coupled model. Part II: Prediction with piggyback data assimilation, *Clim. Dyn.*, *16*, 35–48, 2000.

Tang, Y., R. Kleeman, A. M. Moor, J. Vialard, and A. Weaver, A simple initialization scheme for an oceanic general circulation model for ENSO prediction, *Mon. Wea. Rev.*, submitted, 2003.

Vialard, J., P. Delecluse, and C. Menkes, A modeling study of salinity variability and its effects in the tropical Pacific ocean during the 1993–99 period, *J. Geophys. Res.*, 2000JC000758, 27, September 2002.

Vialard, J., T. A. Weaver, D. L. T. Anderson, and P. Delecluse, Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean, Part 2: physical validation, *Mon. Wea. Rev.*, *131*, 1379–1395, 2003.

Weaver, T. A., J. Vialard, and D. L. T. Anderson, Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean, Part 1: formulation, internal diagnostics and consistency checks, *Mon. Wea. Rev.*, *131*, 1360–1378, 2003.

White, W. B., Design of a global observing system for gyre-scale upper ocean temperature variability, *Prog. in Oceanogr.*, *36*, 169–217, 1995.

A. M. Moore, Program for Atmosphere Ocean Science, University of Colorado, Boulder, USA.

R. Kleeman and Y. Tang, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. (ytang@cims.nyu.edu)

J. Vialard, LODYC, Paris, France.

A. Weaver, Climate Modeling and Global Change Group, C.E.R.F.A.C.S., Toulouse Cedex, France.